

融合内容与用户手势行为的用户画像构建系统设计与实现*

汪强兵 [1,2]　章成志 [1,2,3]
[1](南京理工大学经济管理学院　南京 210094)
[2](江苏省社会公共安全科技协同创新中心　南京 210094)
[3](江苏省数据工程与知识服务重点实验室(南京大学)　南京 210093)

摘要:【**目的**】开发移动端的文献阅读系统, 通过利用在移动端的用户手势行为数据及手势行为对应的内容, 挖掘用户兴趣, 构建用户兴趣画像。【**应用背景**】融合内容与用户行为的用户画像构建系统能够挖掘用户在阅读文献时的兴趣, 并进行用户画像构建。【**方法**】以移动平台下的 Web 阅读系统为工具, 通过收集用户在移动设备上浏览文献产生的用户手势行为(单击、双击、滑动、拖动、放大/缩小等)数据以及与用户手势行为相对应的文本内容, 结合对应文本内容的浏览时间构建用户模型。【**结果**】用户在使用文献阅读系统时可以发现自己在阅读文献过程中的阅读兴趣, 进行用户兴趣画像构建。【**结论**】初步研究结果表明使用用户手势行为可以在一定程度上反映用户的阅读兴趣, 并进行用户建模。该研究结论可以提高市场营销和个性化推荐系统的效果。

关键词: 手势行为　移动设备　文本挖掘　用户建模
分类号: G350

1　引　言

随着经济社会的不断发展, 移动设备越来越成为人们日常生活中不可或缺的工具。随着移动设备的大量普及与功能不断完善, 人们原本只在计算机中进行的操作开始越来越多地转移到移动设备上。计算机上的用户行为已经得到广泛的研究, 利用计算机上的用户行为数据能够很好地进行用户建模, 揭示用户兴趣。由于移动设备的特性, 用户在使用移动设备的过程中会产生不同的手势行为。目前, 利用移动设备的手势行为进行用户建模的研究比较少。本文旨在研究如何使用移动设备中产生的用户手势行为进行用户建模。

当用户在移动设备上浏览一篇文献时, 不可避免地需要在移动设备的屏幕上发生手势行为, 如单击(tap)、上滑(swipe up)、下滑(swipe down)、向上拖动(drag up)、向下拖动(drag down)、放大(pinch out)、缩小(pinch in)等。由于移动设备的使用特点以及屏幕尺寸限制, 当用户在屏幕上执行单击操作时, 表示用户在点击屏幕上的内容; 当用户在屏幕上执行下滑、上滑操作时, 表示正在快速地更换屏幕中显示的内容; 当用户在屏幕中执行向上拖动、向下拖动操作时, 表示用户正在缓慢地更换屏幕中显示的内容, 即用户正在阅读屏幕中的内容; 当用户在屏幕上执行放大操作时, 表示用户正在放大屏幕中的局部内容, 以便更加清晰地阅读。Guo 等[1]的研究表明移动触摸交互(Mobile Touch Interaction), 即用户在移动设备上的手势行为能够反映用户是否对其阅读的文章感兴趣, 而且用户在浏览相关内容的停留时间能够反映用户对此

通讯作者: 章成志, ORCID: 0000-0001-8121-4796, E-mail: zhangcz@njust.edu.cn。
*本文系国家社会科学基金项目"在线社交网络中基于用户的知识组织模式研究"(项目编号: 14BTQ033)的研究成果之一。

内容的兴趣度。

本文提出不同的手势行为在反映用户兴趣方面具有不同的权重，结合手势行为与手势行为对应的内容能够揭示用户兴趣。本文结合自然语言处理技术、Java Web 技术以及可视化技术构建移动阅读平台，用户通过注册之后可以在阅读平台中阅读相关文献。该系统通过收集用户在浏览文献过程中发生的手势行为数据，利用用户在浏览文献过程中产生的 pinch in/out、drag、swipe、tap 手势行为[2]，结合手势行为对应的文本片段以及文本片段浏览时间，发现用户的兴趣并形成用户的兴趣云图。

2 相关研究

随着网络信息数量的快速膨胀，信息过载问题的出现，越来越多的网站开始思考如何不让用户迷失在信息海洋中。通过发现用户兴趣进而向用户推荐其感兴趣的信息或内容是一个很好的解决方法。基于此，很多学者开始研究如何利用用户的相关信息来获取用户兴趣进而实现个性化推荐。Joachim 等[3]研究表明从搜索页面的点击行为数据中能够很好地挖掘出用户的兴趣、偏好信息。孙铁利等[4]根据用户在计算机端 Web 浏览器中的浏览行为建立用户兴趣模型，通过包括拖动滚动条、阅读时间、保存文档、添加书签、打印网页、跟随超链接在内的用户行为确定用户对于一个页面的兴趣度。赵银春等[5]根据用户浏览页面时的内容信息和行为信息，考虑页面的浏览时间和翻页/拉动滚动条的次数，利用多元线性回归模型计算用户对某一页面的兴趣度。Huang 等[6]结合鼠标的点击行为与鼠标在搜索页面不同区域的移动从而提高检索的效果。

上述均是针对计算机端的研究。用户在计算机上的操作主要依赖于键盘以及鼠标等设备，而在移动设备中用户主要依靠手指在触摸屏上不同的手势行为进行相关操作，因此在计算机上的相关研究无法直接运用到移动设备中。随着移动设备的大量普及，学者将注意力转移到研究移动设备中的用户手势行为中。Guo 等[1]比较了移动设备中的触摸操作和计算机设备中使用鼠标和键盘的操作之间的不同，通过挖掘移动设备的用户行为提高检索结果的效果。Han 等[7]利用移动手势行为发现与用户关联度最高的文本片段，进而提高跨设备检索的效果。文献[8]研究表明在用户浏览一篇文章的过程中，并不一定对文章的每一段的内容都感兴趣，而可能只对文章的某一段或者某几段的内容感兴趣。

结合上述研究，本文为不同的手势行为都设定一定的权重，以此来反映用户对发生手势行为文本片段兴趣度的大小。通过统计在文本片段上发生的手势行为的种类与对应频数，结合文本片段的浏览时间发现用户在该文本片段中感兴趣的关键词与对应的兴趣度，进而计算用户在浏览文献过程中感兴趣的关键词。最终，结合用户浏览的文献集合发现用户的兴趣空间，构建用户画像。

3 系统设计

3.1 设计思路

本文首先通过阅读平台收集的手势行为数据与阅读时间信息确定用户在浏览文章过程中对不同文本片段的兴趣度；其次，综合文章中所有发生手势行为的文本片段确定用户在浏览文章时感兴趣的关键词；最终结合用户浏览的所有文章发现用户的兴趣，构建用户模型并可视化处理。设计思路如图 1 所示。

3.2 系统架构设计

根据上述的设计思路，将系统的架构分为三层，如图 2 所示。

(1) 数据层

数据层负责保存系统运行时所需要的数据、用户在浏览文献过程中产生的行为数据以及用户注册时填写的个人信息。

(2) 处理层

处理层负责系统的数据处理任务，包括手势搜集、中文分词、关键词抽取、兴趣度计算、画像构建以及数据可视化任务。通过手势搜集将用户在浏览文献过程中发生的手势行为数据存入数据库中；利用中文分词、关键词抽取等自然语言处理技术得到用户在阅读时的兴趣词；根据得到的兴趣词刻画用户兴趣画像并使用数据可视化工具展示用户画像。

(3) 视图层

视图层包括用户个人信息的管理、文献阅读以及用户兴趣画像的可视化展示。

3.3 关键技术描述

(1) 文本片段兴趣度的计算

本文通过分析用户浏览文献过程中的手势行为与

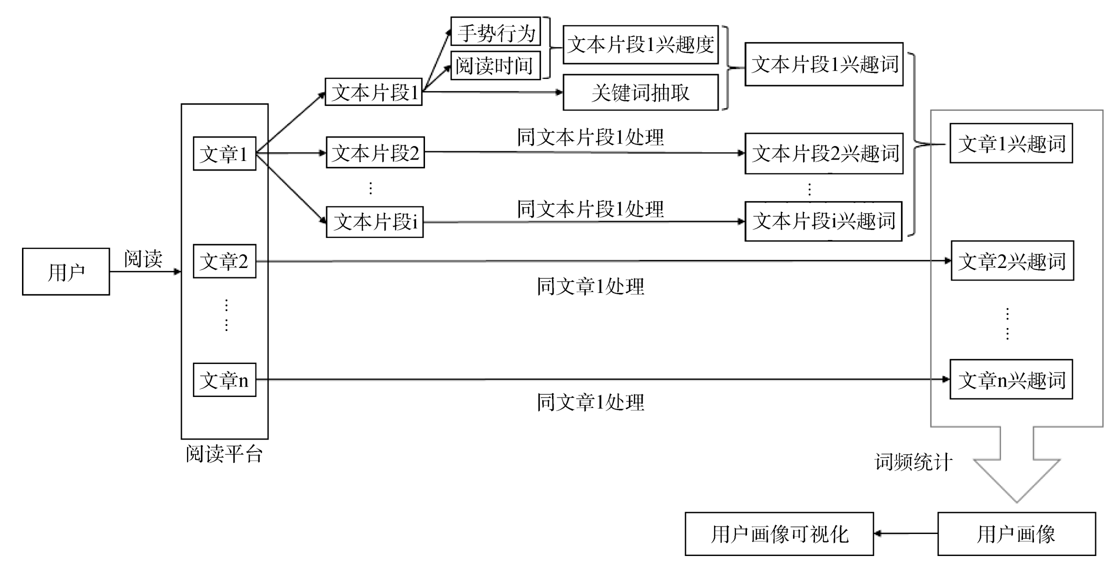


图 1 设计思路

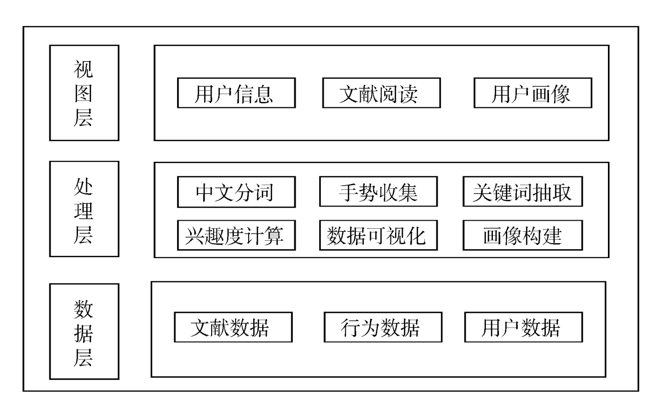


图 2 系统架构

浏览时间获取用户感兴趣的文本片段，并计算文本片段的兴趣度。确定了不同手势行为，反映用户对文本片段感兴趣度的权重大小。pinch in 手势行为表示用户在缩小屏幕中的内容；pinch out 表示用户在放大屏幕中的内容；drag 表示用户手指在屏幕上缓慢地滑动；swipe 表示用户手指在屏幕上快速地滑动。由于 swipe 手势行为表示用户正在快速地更换屏幕中显示的内容，所以 swipe 对反映文本片段兴趣度的贡献值很小[2]，据此本文将 swipe 的权重设置为 0。对 pinch in/out、drag、tap 采用层次分析法确定各自的权重。在用户浏览文章时发生的手势行为中，pinch in/out 操作的权重最高，其次是 drag 操作，而 tap 操作的权重最小。据此构建判断矩阵，如表 1 所示。

表 1 中，第二行第三列中的 3 代表 pinch in/out 行为比 drag 行为稍微重要/有优势，第二行第四列中的 5 代表 pinch in/out 行为比 tap 行为比较重要/有优势。经过计算得出 pinch in/out 的权重为 0.6267，drag 的权重为 0.2797，tap 的权重为 0.0936。一致性检验结果为 0.0825，计算结果有效。至此得到手势行为对于文本片段兴趣度的权重，如表 2 所示。

表 1 判断矩阵

手势行为	pinch in/out	drag	tap
pinch in/out	1	3	5
drag	1/3	1	4
tap	1/5	1/4	1

表 2 手势行为权重

手势行为	pinch in/out	drag	tap	swipe
权重	0.6267	0.2797	0.0936	0

文献由不同的文本片段组成，用户在不同的文本片段会产生不同的浏览时间。本文利用文本片段上发生的手势行为测量用户在该文本片段的浏览时间。即在用户浏览文献时，系统后台会自动记录下用户手势行为第一次作用于文本片段的时间。本文把手势行为第一次作用于该文本片段的时间与用户离开该文本片段后手势行为第一次作用于其他文本片段的时间差作为该文本片段的浏览时间。计算文本片段 P_i 的浏览时间 t_i 如公式(1)所示。

$$t_i = t_i^f - t_j^f \quad (1)$$

其中，t_i^f 表示文本片段 P_i 第一次发生手势行为的

时刻，t_j^f 表示离开文本片段 P_i 后手势行为第一次发生在其他文本片段上的时刻。

用户在文章中不同文本片段的浏览时间之和即为用户浏览该文章的时间。用户浏览文章的时间 T 的计算方法如公式(2)所示。

$$T=\sum_{i=1}^{n}t_i \quad (2)$$

其中，t_i 表示文本片段 P_i 的浏览时间，n 表示文章中发生手势行为的文本片段个数。

文本片段的浏览时间能够反映出用户对文本片段的兴趣度；同时，发生在文本片段上的手势行为同样能够反映用户对文本片段的兴趣度。本文结合时间与手势行为信息，综合计算用户对文本片段的兴趣度。文本片段 i 中手势行为 pinch in/out、drag、tap、swipe 发生的次数分别为 f_p、f_d、f_t、f_s，对应的权重分别为 v_p、v_d、v_t、v_s，文本片段 P_i 的浏览时间为 t_i，文本片段所属文章的总浏览时间为 T。设文本片段 P_i 对应的权重为 $W(i)$，则文本片段 P_i 的兴趣度计算方式如公式(3)所示。

$$W(i)=\frac{t_i}{T}\times(f_p\times v_p+f_d\times v_d+f_t\times v_t+f_s\times v_s) \quad (3)$$

(2) 文本片段的关键词抽取及向量空间表示

本文采用 ICTCLAS①进行文本片段关键词的抽取。ICTCLAS 是一款处理中文文本的程序包，它可以完成文本分词、计算关键词、发现新词等文本处理任务。ICTCLAS 基于信息熵原理提取文本中的关键词。提取出关键词后保留关键词的权重并将权重归一化处理，处理过程如公式(4)所示。

$$NC_i=\frac{C_i-C_{\min}}{C_{\max}-C_{\min}} \quad (4)$$

其中，C_i 表示文本中利用 ICTCLAS 处理后某个词的关键词权重，$C_{\max}$ 是文本中全部关键词权重的最大值，$C_{\min}$ 是文本中全部关键词权重的最小值，NC_i 为归一化处理后关键词的权重。

综上，笔者对发生手势行为的文本片段抽取关键词之后，得到一组用来表示该文本片段的关键词序列以及与关键词对应的权重。通过使用向量空间模型来表达该文本片段，文本片段 P_i 表达方式如下所示:

$$P_i=\{(K_1, NC_1), (K_2, NC_2), \cdots, (K_n, NC_n)\}$$

其中，K_n 表示 P_i 段中抽取的第 n 个关键词。

(3) 用户画像生成及可视化

在完成手势行为与浏览时间信息得到文本片段兴趣度和利用关键词抽取技术获得文本片段关键词之后，计算用户的兴趣关键词，构建用户的兴趣空间。由于用户在浏览特定文献时，并不是对文献中的所有内容都感兴趣，而是关注于文献中的某几个文本片段。把发生手势行为的文本片段提取出来，重新组成用户浏览的文献 D_i，即:

$$D_i=(P_1, P_2, \cdots, P_n)$$

其中，P_n 表示文献 D_i 中发生了手势行为的文本片段。

已经获得文本片段 P_i 的兴趣度 $W(i)$ 与关键词权重 NC_i。笔者将其权重均设置为 0.5。根据文本片段兴趣度计算 P_i 中关键词 K 的兴趣度。计算方法如公式(5)所示。

$$\text{关键词 } K \text{ 兴趣度}=0.5\times NC_i+0.5\times W(i) \quad (5)$$

其中，K 表示文本片段 P_i 中抽取的关键词。

通过上述步骤，获得文章 D_i 中每一个发生手势行为的文本片段中的关键词兴趣度。在得到文本片段中关键词兴趣度之后，进一步计算用户在文章 D_i 中最感兴趣的关键词。如果关键词在不同发生手势行为的文本片段中出现，计算其平均兴趣度，以此作为用户在浏览文章 D_i 时对该关键词的兴趣度。计算方法如公式(6)所示。

$$KI=\frac{\sum_{i=1}^{n}P_i(NC)}{n} \quad (6)$$

其中，$P_i(NC)$ 表示关键词在文本片段 P_i 中的兴趣度，n 为文章 D_i 中包含某一个关键词的个数。分子表示关键词在不同文本片段中的兴趣度之和。

综上即得到文章 D_i 用户感兴趣的关键词以及对应的兴趣度，选择兴趣度最大的 10 个关键词用来表示用户在浏览该文章过程中最感兴趣的关键词。使用同样的方法计算用户浏览的所有文章，每一篇文章都包含用户兴趣度最高的 10 个关键词。最终将处理后的文章组成用户浏览的文章集合 $D=(D_1, D_2, \cdots, D_n)$。在文章集合 D 中，分别计算每一个关键词在文档集合中出现的频率，进而得到用户的兴趣空间。用户的兴趣画

①http://ictclas.nlpir.org/.

像表示为如下集合:

$$\{(k_1, m_1), (k_2, m_2), \cdots, (k_i, m_n)\}$$

其中, m_i 表示关键词出现的频率, i=1, 2, ⋯, n。

在得到用户画像后, 笔者使用数据可视化工具更加清晰地展示用户画像。本文采用百度公司提供的可视化工具 ECharts①实现用户画像的可视化工作。通过使用 ECharts 中的字符云图展示用户画像, 其中用户画像中的关键词表示字符云图上的字符, 对应的频率作为字符的权重用来控制字符的字体大小。

4 系统开发与应用

4.1 系统开发

为了收集用户在浏览过程中的手势行为, 开发了手机阅读网站——文献阅读系统, 其界面如图 3 所示。用户通过注册文献阅读系统, 便可以使用相应的账号登录系统并阅读相关文献, 在阅读过程中, 用户可以进行评论、转发、点赞等行为, 而本文主要研究通过用户在浏览页面过程中发生的手势行为信息结合浏览时间来挖掘用户感兴趣的领域。



(a) 登录页面 (b) 标题页面 (c) 内容页面

图 3 系统展示

系统可以识别用户在浏览页面过程中发生的放大(pinch out)/缩小(pinch in)、滑动(swipe)、拖动(drag)、点击(tap)手势行为, 并记录手势行为对应的文本片段文本及文本片段浏览时间。本文采用 JavaScript 开源工具包②识别用户的手势行为。在设计收集手势行为对应的文本片段时参考了 Han[9]的相关 js 代码。

系统设计完毕后, 用户可以注册、登录系统, 然后根据自身的兴趣爱好和研究领域阅读相关文献。

4.2 系统应用

为检测该系统的可行性, 招募研究组中的一个研究生进行相关实验。实验之前, 介绍了系统基本情况及系统的使用方法; 被试人员根据自己的兴趣爱好浏览系统的相关文献。

系统将前台收集到的用户手势数据、对应的文本片段文本以及其他相关数据存入数据库中, 数据库中的信息包括用户 ID(userID)、阅读的文章 ID (documentID)、手势行为发生的时间(time)发生在文本片段文本上的手势行为(type)、文本片段文本(touchHtml)以及手势行为的详细数据(手势行为发生的位置、手势行为的速度、时间戳等)。用户 ID 为 user1 浏览第 32 篇文章的手势行为序列如表 3 所示(文本片段数据较多, 省略显示)。

表 3 手势行为序列

userID	documentID	time	type	touchHtml
user1	32	2016-01-10 13: 47: 01	drag	paragraph1
user1	32	2016-01-10 13: 47: 02	drag	paragraph1
user1	32	2016-01-10 13: 47: 13	drag	paragraph2
user1	32	2016-01-10 13: 47: 14	drag	paragraph2
user1	32	2016-01-10 13: 47: 19	drag	paragraph3
user1	32	2016-01-10 13: 47: 20	drag	paragraph3
user1	32	2016-01-10 13: 47: 25	drag	paragraph3
user1	32	2016-01-10 13: 47: 55	tap	paragraph4
user1	32	2016-01-10 13: 48: 17	drag	paragraph5
user1	32	2016-01-10 13: 48: 18	drag	paragraph5
user1	32	2016-01-10 13: 48: 19	drag	paragraph5
user1	32	2016-01-10 13: 48: 20	drag	paragraph5
user1	32	2016-01-10 13: 48: 39	drag	paragraph6
user1	32	2016-01-10 13: 48: 40	drag	paragraph6
user1	32	2016-01-10 13: 48: 41	drag	paragraph6
user1	32	2016-01-10 13: 48: 42	drag	paragraph6
user1	32	2016-01-10 13: 48: 43	drag	paragraph6
user1	32	2016-01-10 13: 48: 45	drag	paragraph6

(注: paragraph 后的数字 i 并不代表属于文章中的第 i 段, 而只是为了区别不同的文本片段。)

①http://echarts.baidu.com.cn/.

②http://hammerjs.github.io/.

在剔除无效数据之后，经过处理获得用户在阅读一篇文章时的手势行为序列与对应的文本片段文本的时间信息。

采取上文计算文本片段的浏览时间的方法得到文本片段的浏览时间，其次计算文片段上发生手势行为的种类与对应的次数，readtime 表示该文本片段阅读的时间，dragtime 表示 drag 操作在该文本片段中发生的次数，swipetime 表示 swipe 操作在该文本片段发生的次数，taptime 表示 tap 操作发生的次数，pinchintime 表示 pinch in 操作发生的次数，pinchouttime 表示 pinch out 操作发生的次数，得到的结果如表 4 所示。最终得到 user1 在浏览文章 D_i 过程中感兴趣的关键词以及对应的兴趣度。

表 4　手势行为序列处理结果

userID	documentID	readtime	dragtime	swipetime	taptime	pinchintime	pinchouttime	touchHtml
user1	32	0 分 12 秒	1	0	0	0	0	paragraph1
user1	32	0 分 6 秒	2	0	0	0	0	paragraph2
user1	32	0 分 36 秒	2	0	0	0	0	paragraph3
user1	32	0 分 22 秒	0	0	1	0	0	paragraph4
user1	32	0 分 22 秒	2	0	0	0	0	paragraph5
user1	32	0 分 6 秒	4	0	0	0	0	paragraph6

使用同样方法处理 user1 阅读的其他文章，形成用户浏览的文章集合。根据文章集合得到的用户兴趣空间如下：

I_{user1}={(政府知识管理, 2), (个人信息管理, 2), (社会化, 1), (UTAUT, 1), (系统, 1), (信任, 1), (信任信念, 1), (政府, 1), (专利审核, 1), (模型 , 1), (问题, 1), (信息, 1), (智能手机, 1), (问题解决, 1), (个人信息, 1), (感知有用性, 2), (社会化媒体, 1), (技术采纳, 1), (微博, 2), (社会化媒体应用, 3)}

采用 ECharts 处理上述的用户兴趣空间。云图中，用户对某一个关键词的兴趣越大，该关键词的字体越大。结果如图 6 所示。

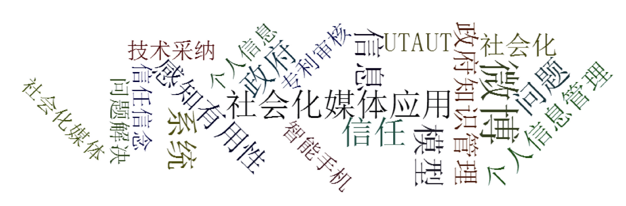


图 4　用户兴趣云图

5　结　语

本文结合 Java Web 相关技术构建了移动阅读系统。该系统可以捕捉用户在阅读时发生的手势行为。用户注册以后，即可通过移动设备登录系统并在系统上阅读自己感兴趣的文献。在用户阅读过程中，系统会自动记录下用户在阅读时发生的手势行为、手势行为发生的时间以及发生手势行为的文本片段并将数据存入服务器中。在获得用户相关的手势行为数据后，系统使用自然语言处理技术构建用户兴趣画像，并最终使用可视化技术展示用户兴趣画像。

本文没有比较利用移动设备中的用户手势行为建模与使用传统方式建模之间的差异，在今后的工作中将进行相关实验。用户在浏览文章过程中，除了发生手势行为，还会发生其他行为，如对一篇文章点赞、转发文章、评论文章等。本文在建立用户兴趣空间时没有加入点赞、转发、评论等影响因素，在今后的工作中，将结合这些因素更加深入地研究用户兴趣爱好。

参考文献：

[1] Guo Q, Jin H, Lagun D, et al. Mining Touch Interaction Data on Mobile Devices to Predict Web Search Result Relevance [C]//Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2013: 153-162.

[2] Han S G, Hsiao I H, Parra D. A Study of Mobile Information Exploration with Multi-Touch Interactions[C]//Proceedings of the 7th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. 2014: 269-276.

[3] Joachims T, Granka L, Pan B, et al. Accurately Interpreting Clickthough Data as Implicit Feedback[C]//Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2005: 154-161.

[4] 孙铁利, 杨凤芹. 根据用户隐式反馈建立和更新用户兴趣模型[J]. 东北师大学报: 自然科学版, 2003, 35(3): 99-104.

(Sun Tieli, Yang Fengqin. An Approach of Building and Updating User Interest Profile According to the Implicit Feedback [J]. Journal of Northeast Normal University: Natural Science Edition, 2003, 35(3): 99-104.)

[5] 赵银春, 付关友, 朱征宇. 基于 Web 浏览内容和行为相结合的用户兴趣挖掘[J]. 计算机工程, 2005, 31(12): 93-94. (Zhao Yinchun, Fu Guanyou, Zhu Zhengyu. User Interest Mining of Combining Web Content and Behavior Analysis [J]. Computer Engineering, 2005, 31(12): 93-94.)

[6] Huang J, White R W, Dumais S T. No Clicks, No Problem: Using Cursor Movements to Understand and Improve Search [C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011: 1225-1234.

[7] Han S G, Yue Z, He D Q. Understanding and Supporting Cross-Device Web Search for Exploratory Tasks with Mobile Touch Interactions [J]. ACM Transactions on Information Systems, 2015, 33(4): Article No. 16.

[8] Morita M, Shinoda Y. Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval[C]// Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 1994: 272-281.

[9] Han S. PITT (PIck-up The Touches) -- A Javascript Plugin to Track Your Website Visits [EB/OL]. [2016-06-25]. http://www.pitt.edu/~shh69/pitt.html.

作者贡献声明:

章成志: 提出研究思路, 设计研究方案, 论文最终版本修订;
汪强兵: 处理数据, 起草论文。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: 1906439961@qq.com, zhangcz@njust.edu.cn。

[1] 汪强兵, 章成志.userdata.mdb.用户手势行为数据.

收稿日期: 2016-08-25
收修改稿日期: 2016-11-07

Constructing Users Profiles with Content and Gesture Behaviors

Wang Qiangbing[1, 2] Zhang Chengzhi[1, 2, 3]
[1](School of Economics and Management, Nanjing University of Science & Technology, Nanjing 210094, China)
[2](Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094, China)
[3](Jiangsu Key Laboratory of Data Engineering and Knowledge Service (Nanjing University), Nanjing 210093, China)

Abstract: [**Objective**] This paper constructs users profiles by gauging their interests from gesture behaviors and related contents from a mobile article reading system. [**Context**] Users profiles construction with content and gesture behaviors can identifies users' mobile reading interests and profiles effectively. [**Methods**] First, we collected user gesture behaviors (such as tap, double tap, swipe, drag, pinch in/out) as well as corresponding contents from a mobile article reading system. Second, we established the users model based on the collected data and reading time. [**Results**] Users could find their own reading interests while browsing papers with our system, which help us build users profiles. [**Conclusions**] Users gesture behaviors reveal their reading interests, which could improve the performance of marketing and personalized recommendation systems.

Keywords: Gesture Behaviors Mobile Device Text Mining User Modeling